# Phonon-accurate machine-learning potentials from automated workflows

BAM
Bundesanstalt für Materialforschung und -prüfung

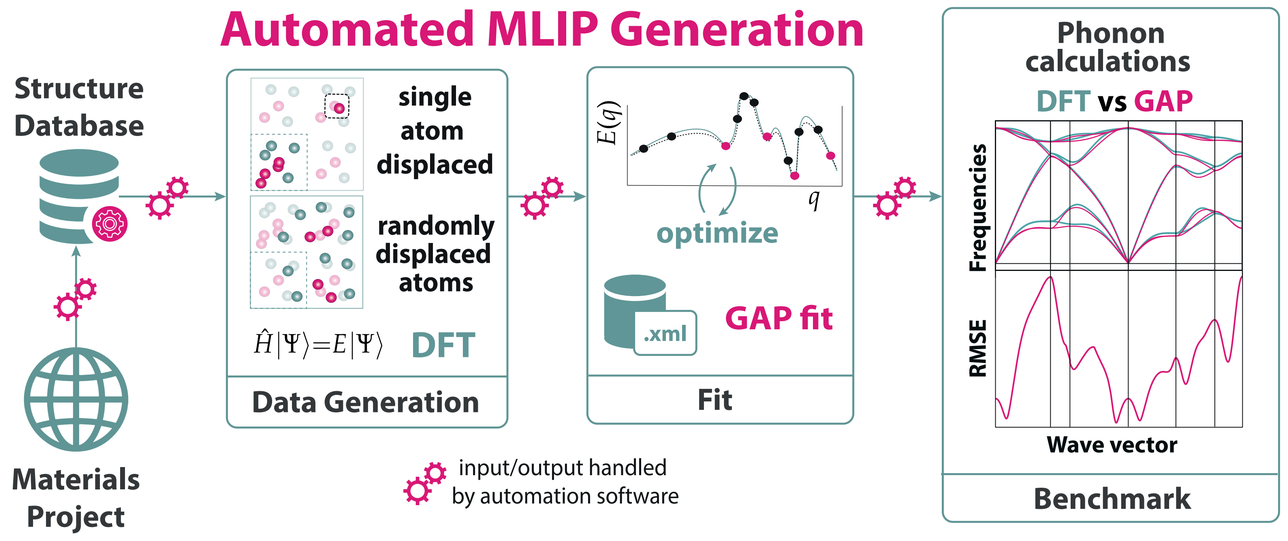
**Christina Ertural[1*], Volker L. Deringer[2], Janine George[1,3] (email: christina.ertural@bam.de)**

[1]Federal Institute for Materials, Research and Testing, Department Materials Chemistry, Unter den Eichen 87, 12205 Berlin, Germany.
[2]Inorganic Chemistry Laboratory, Department of Chemistry, University of Oxford, Oxford OX1 3QR, United Kingdom
[3]Friedrich Schiller University Jena, Institute of Condensed Matter Theory and Solid-State Optics, Max-Wien-Platz 1, 07743 Jena, Germany.

## Introduction

Data-driven materials design aims to predict and optimize material properties, such as stability and thermal conductivity, which are influenced by vibrational behavior. Approaches like DFT are computationally demanding and have limitations for phonon calculations. Machine learning-driven interatomic potentials (MLIP), like the Gaussian approximation potential (GAP), offer a more efficient alternative.[1–8]

We developed a Python workflow to automate MLIP generation using the Materials Project database.[9] DFT computations, MLIP fitting and benchmark steps are automated.[10,11] This approach accelerates phonon calculations and supports testing different data generation strategies and hyperparameters, and further validation[12] is planned. Our goal is to provide open-source code and share these potentials to enhance reproducibility and accessibility in computational chemistry.




## Results

The Python workflow was used to automatically construct a Si database (of single-atom and randomly displaced supercells) from 12 Si allotropes and perform SOAP-only GAP fits iterating through various combinations of convergence parameter and hyperparameter sets (n_sparse, SOAP delta). We present the results for MPIDs mp-149, mp-1095269, and mp-971662. All data is presented for the atomwise regularization parameter of 0.1.

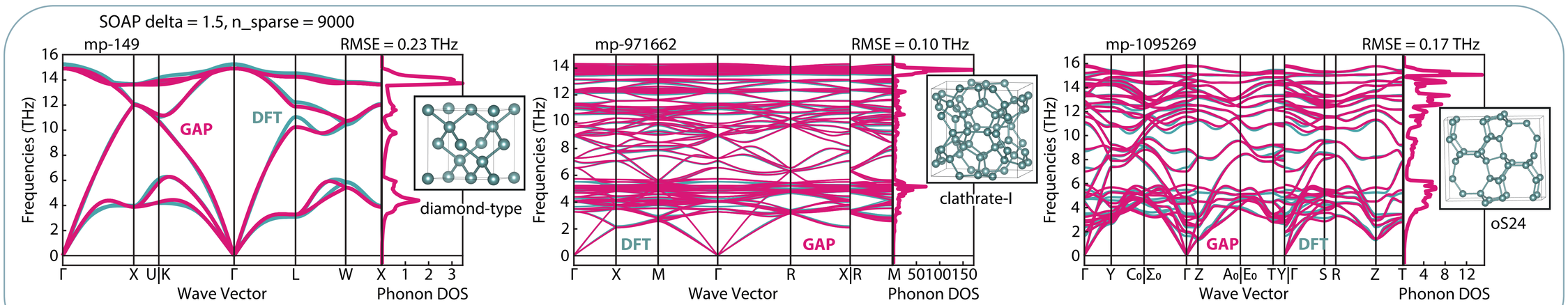



Fig. 1: Phonon bandstructure comparison and DOS for mp-149, mp-971662 and mp-1095269. The insets show the unit cell structure.

retrieve all phonon RMSE for each parameter setup

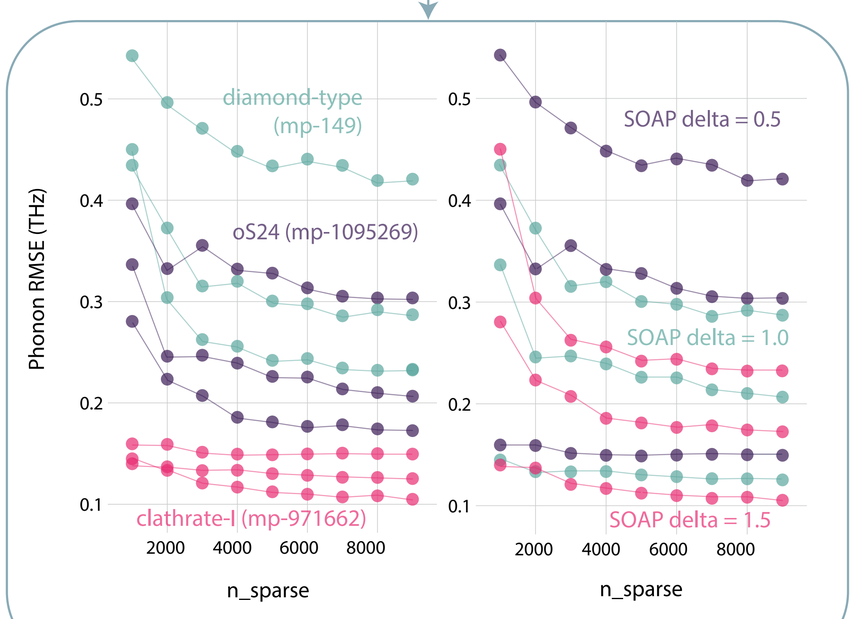



Fig. 2: Phonon RMSE vs. n_sparse.

retrieve all phonon RMSE for each force test
RMSE in dependence of SOAP delta and n_sparse

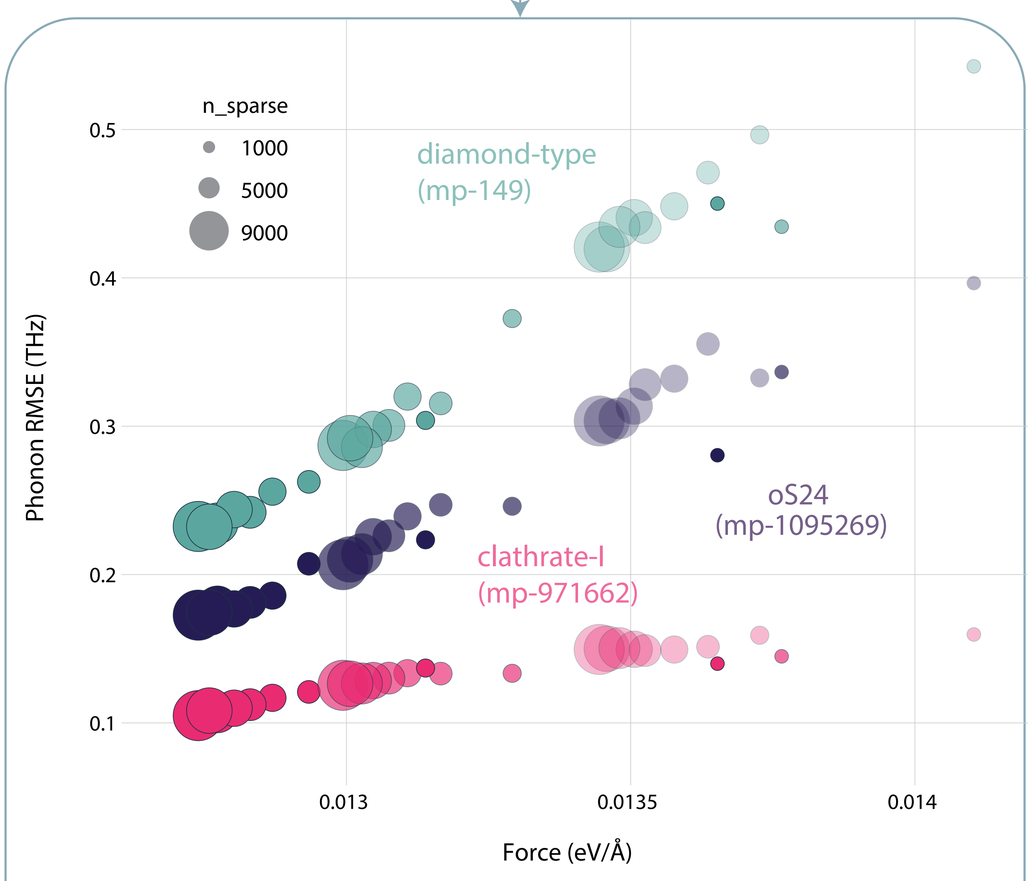



Fig. 3: Phonon RMSE vs. force RMSE for the test data. The n_sparse values are used for resizing the data points and the SOAP delta (0.5, 1.0, 1.5) for different intensities.

## Summary and Outlook

>automated MLIP fit and benchmark successful
>overall good correlation between different error metrics
>quality of the fits and of the results improve a lot with an increasing n_sparse value
>results for clathrate-I type Si are best overall
>the best-performing parameter setup (SOAP delta = 1.5, n_sparse = 9000) generates comparibly good results for the three shown Si allotropes

>improve data generation and data quality [WIP]
>more GAP fits with other atomwise regularization parameters [WIP] and repeat fit and data analysis for individual data types (single-atom/randomly displaced) [WIP]
>check thermal properties (it is known from the literature[7] that a too tight fit, i.e. a small atomwise regularization parameter, leads to instabilities in calculating thermal properties)
>test other and more diverse chemical systems (e.g. $Sb_2Se_3$) [WIP]
>test different MLIP architectures (framewroks are implemented in the Python workflow already)

(1) Snyder, G. J.; Toberer, E. S. Nat. Mater. 2008, 7 (2), 105–114. (2) He, J.; Xia, Y.; Lin, W.; Pal, K.; Zhu, Y.; Kanatzidis, M. G.; Wolverton, C. Adv. Funct. Mater. 2022, 32 (14) 2108532. (3) Wehmeyer, G.; Yabuki, T.; Monachon, C.; Wu, J.; Dames, C. Appl. Phys. Rev. 2017, 4 (4), 041304. (4) Togo, A.; Tanaka, I. Scr. Mater. 2015, 108, 1–5. (5) Hellman, O.; Steneteg, P.; Abrikosov, I. A.; Simak, S. I. Phys. Rev. B 2013, 87 (10), 104111. (6) Eriksson, F.; Fransson, E.; Erhart, P. Adv. Theory Simul. 2019, 2 (5), (7) George, J.; Hautier, G.; Bartók, A. P.; Csányi, G.; Deringer, V. L. J. Chem. Phys. 2020, 153 (4), 044104. (8) Bartók, A. P.; Payne, M. C.; Kondor, R.; Csányi, G. Phys. Rev. Lett. 2010, 104 (13), 136403. (9) Jain, A.; Ong, S. P.; Hautier, G.; Chen, W.; Richards, W. D.; Dacek, S.; Cholia, S.; Gunter, D.; Skinner, D.; Ceder, G.; Persson, K. A. APL Mater. 2013, 1, 011002. (10) Mathew, K.; Montoya, J. H.; Faghaninia, A.; Dwarakanath, S.; Aykol, M.; Tang, H.; Chu, I.; Smidt, T.; Bocklund, B.; Horton, M.; Dagdelen, J.; Wood, B.; Liu, Z.-K.; Neaton, J.; Ong, S. P.; Persson, K.; Jain, A. Comput. Mater. Sci. 2017, 139, 140–152. (11) Ong, S. P.; Richards, W. D.; Jain, A.; Hautier, G.; Kocher, M.; Cholia, S.; Gunter, D.; Chevrier, V. L.; Persson, K. A.; Ceder, G. Comput. Mater. Sci. 2013, 68, 314–319. (12) Morrow, J. D.; Gardner, J. L. A.; Deringer, V. L. J. Chem. Phys. 2023, 158 (12), 121501.

Sicherheit in Technik und Chemie

How to reach us: jageo.github.io quantumchemist.github.io

www.bam.de

FOCUS AREA MATERIALS